Operational Experience with the ALICE High Level Trigger

Artur Szostak[1] for the ALICE collaboration

[1] Department of Physics and Technology, University of Bergen, Norway

Email: artursz@iafrica.com

Introduction

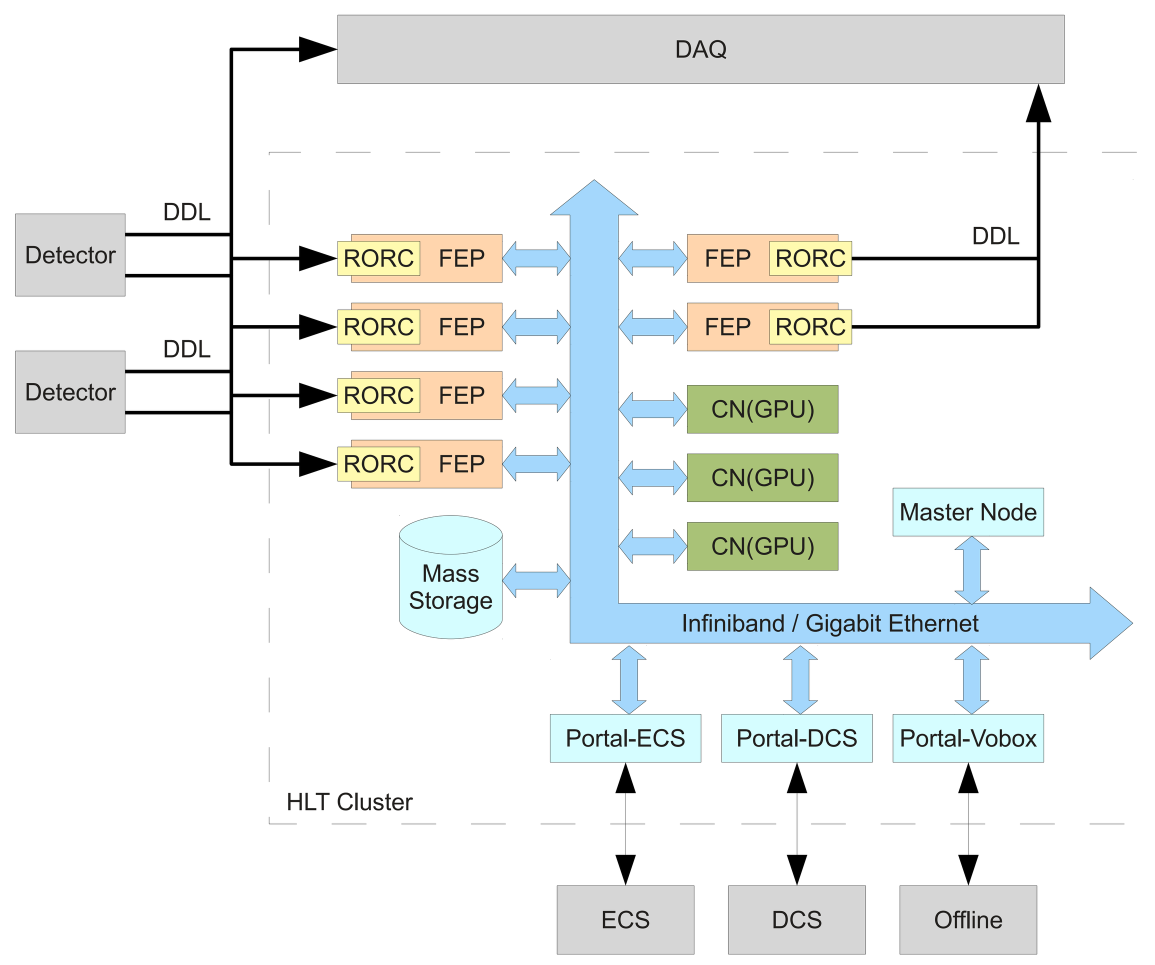
The ALICE High Level Trigger (HLT) is a dedicated real-time system for online event reconstruction and triggering [1]. Its main goal is to reduce the 25 GB/s of raw data read out from the detectors to fit within the available data acquisition bandwidth of 4 GB/s. This is accomplished by a combination of data compression and triggering. Only when a reconstructed event is selected by HLT is the data recorded. The combination of both approaches allows for flexible data reduction strategies. A second vital function is online monitoring [2]. The HLT has access to all raw data and status information from the detectors during data taking. Combined with online event reconstruction the HLT becomes a powerful monitoring tool for ensuring data quality. The HLT has become vital to the data taking for ALICE during the 2011 Pb-Pb period, when it provided data compression for the Time Projection Chamber (TPC) detector [3]; which contributes over 85% to the total data volume. This has allowed to record 160 million (77% more than compared to 2010) ion collisions during the 2011.

Architecture

Event reconstruction requires a large amount of computing power; provided by a large computing cluster, comprising 248 machines built from commodity components. The machines are divided into the following classes:

- 117 Front End Processor (FEP) machines – contain custom Readout Receiver Cards (RORCs) used to connect to detectors and data acquisition (DAQ).
- 20 Compute Nodes (CN) – executes the general parts of the HLT calculation.
- 64 GPU equipped compute nodes (CNGPU) – used for particle tracking [4].
- 4 portal machines – interfaces the HLT with external subsystems and provides master node functionality.
- 20 infrastructure servers – provide the core cluster services like mass storage.
- 23 development machines – provide a software development and testing environment.

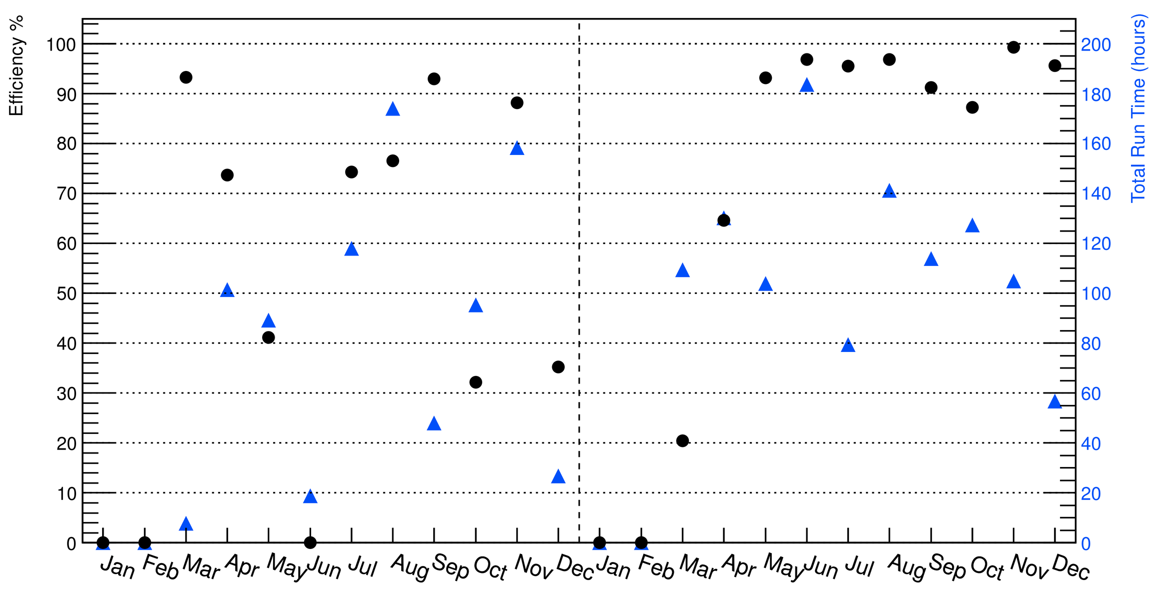
In total there are 2744 cores in the production cluster, with about 2 GB of memory per core (totalling 5.29 TB distributed memory). All machines are interconnected with an InfiniBand network and Gigabit Ethernet for management.



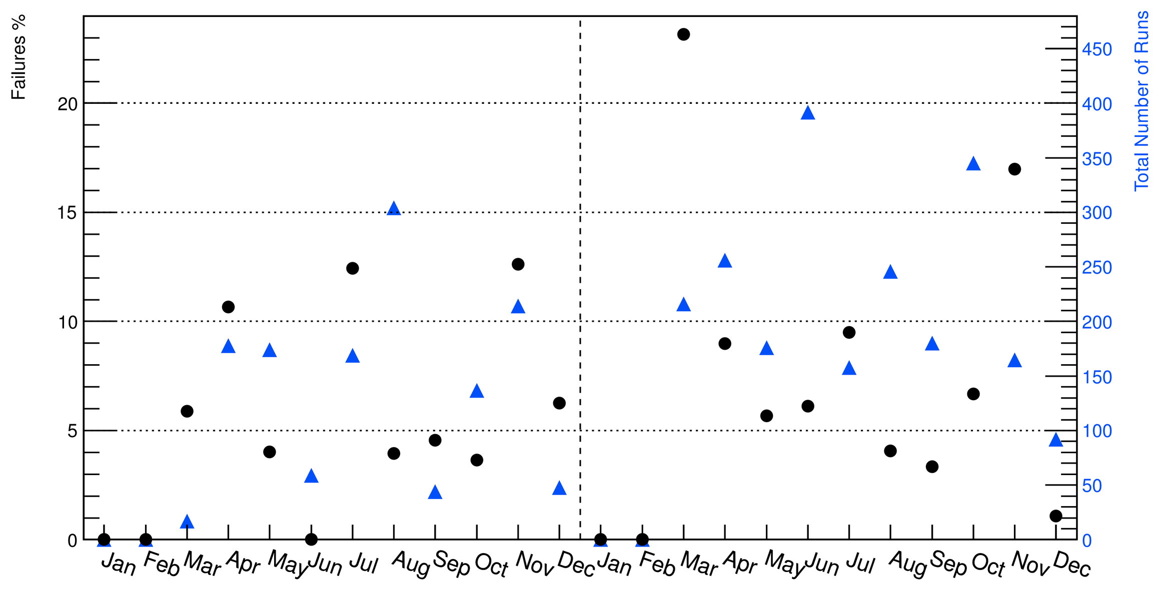
HLT's physical architecture schematic.

Measurements

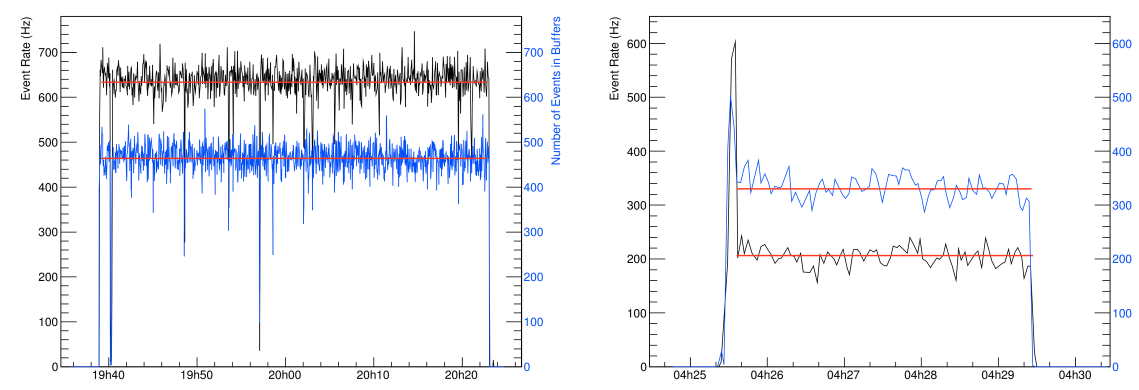
Operational performance has been estimated by measuring the efficiency and failure fractions. In addition, the boundary conditions for maximum event rates and input/output data rates were measured during two special runs (minimum bias and central) taken during the 2011 Pb-Pb period.



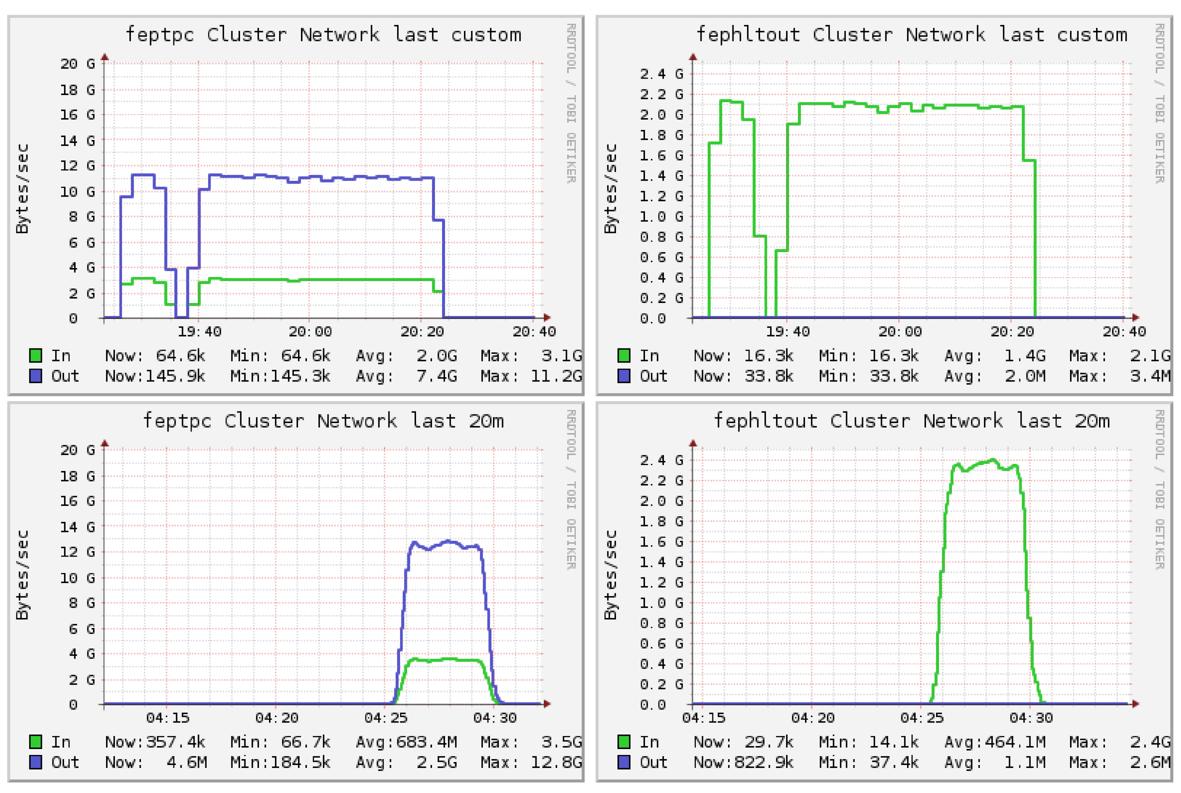
Monthly HLT operational efficiency for 2010-2011.



Monthly HLT failure fraction for 2010-2011.



Event rates and events in buffers for minbias run (left) and central run (right).



Input and output rates for the minimum bias run (top) and central run (bottom). Input rates are indicated by the difference between output and input rates on TPC FEP machines. Output rates are indicated by input rates into FEPs connected to DAQ.

Integrated operational efficiencies and failure rates.

Year	pp Period		Pb-Pb Period	
	Efficiency %	Failures %	Efficiency %	Failures %
2010	63.6	6.2	80.5	11.5
2011	81.8	7.4	98.0	11.3

Performance parameters measured for HLT during the 2011 Pb-Pb period.

Run Type	Event Rate (Hz)	Number of Events in Buffers	Latency (s)	Data Rate (GB/s)	
				Input	Output
Minbias	633.4 ± 2.3	464.0 ± 1.1	0.733 ± 0.003	8.0	2.1
10% Central	201.6 ± 2.8	330.0 ± 4.6	1.64 ± 0.03	9.2	2.3

Modifications

Important modifications were made to the system during 2011 to significantly improve the performance and stability. Both the mass storage system and network were restructured. The new network layout has insulated the production cluster from the development/testing machines. The file system was changed from the Andrew File System (AFS) to Fraunhofer File System (FhGFS), which better suits the performance needs of HLT. Adding additional mass storage nodes and improving the software distribution system have allowed HLT start up times to be reduced from 120-140 s during 2010 down to 50-60 s in 2011 [5]. In addition, special error monitoring agents have allowed to identify problems much sooner leading to improved efficiency.

Conclusion

The HLT has proven a valuable tool for data reduction that can handle all rates seen during 2010-2011. However, it is a complex system that interfaces with many other subsystems in ALICE. 2010 was used as a commissioning period to get the key features of HLT running. The major restructuring work has shown to be effective in significantly improving the operational efficiency. Thus, valuable lessons were learned. The current system is in much better shape, but still some lingering problems need addressing in future upgrades. Better handling of online calibration for reconstruction and improved testing/commissioning procedures for ALICE as a whole are required, to spot difficult bugs only seen when multiple subsystems interact.

References

[1] ALICE Collaboration 2004 *Technical Design Report of the Trigger, Data Acquisition, High-Level Trigger and Control System* CERN/LHCC 2003-062

[2] Hege Erdal et al. 2012 Monitoring the data quality of the real-time event reconstruction in the ALICE High Level Trigger *JoP: Conference Series*, CHEP

[3] ALICE Collaboration 1999 *Technical Design Report of the Time Projection Chamber* CERN/LHCC 2000-001

[4] David Rohr et al. 2012 ALICE HLT TPC Tracking of Pb-Pb Events on GPUs *JoP: Conference Series*, CHEP

[5] Dinesh Ram et al. 2012 Flexible event reconstruction software chains with the ALICE High-Level Trigger *JoP: Conference Series*, CHEP